AD733688

AD

SA GROUP TECHNICAL PAPER TP 1-72

UNITED STATES ARMY COMBAT DEVELOPMENTS COMMAND

FORT BELVOIR, VIRGINIA 22060

The Role of Design in Experimental Investigations

JANUARY 1972

APPROVED FOR PUBLIC RELEASE - DISTRIBUTION UNLIMITED

6- 83,503—Ft Belvoir

26 R

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Systems Analysis Group U.S. Army Combat Developments Command	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

The Role of Design in Experimental Investigations

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Carl B. Bates

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
January 1972	25	10

8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) TP 1-72 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for Public Release - Distribution Unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

The three basic principles of experimental design, replication, randomization, and local control, are briefly discussed. Advantages of the proper application of the principles are given, and the importance of the role of experimental design is illustrated by a numerical example. Three analyses are performed on the same set of test data. First, an analysis is performed on a set of empirical data to salvage as much information as possible from the data of the uncontrolled experiment. Secondly, some (but not all) information is obtained about the conduct of the experiment, necessary assumptions are then made, and an analysis of variance is performed. Finally, the data are properly analyzed by performing a nested-factorial analysis of variance as dictated by the design and the actual conduct of the experiment. The validity and the amount of information resulting from the three analyses are then compared.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Experimental design						
Statistical inference						
Estimation						
Hypotheses testing						
Analysis of variance						

Technical Paper TP-1-72
January 1972

Systems Analysis Group
US Army Combat Developments Command
Fort Belvoir, Virginia 22060

THE ROLE OF DESIGN IN EXPERIMENTAL INVESTIGATIONS

by

Carl B. Bates

ACN 14521

Approved by

PHILIP A. D. SCHNEIDER, Ph.D
Director of Scientific Services

MARION R. BRYSON, Ph.D
Technical Director

HERBERT J. CHILDRESS, JR.
Colonel, ADA
Commanding Officer

TABLE OF CONTENTS

		Page
ABSTRACT		ii
I.	INTRODUCTION	1
II.	BASIC PRINCIPLES OF EXPERIMENTAL DESIGN	2
III.	"SALVAGE" OPERATIONS ON AN UNCONTROLLED EXPERIMENT	5
IV.	"DESIGNED" AFTER EXECUTION	10
V.	PROPERLY DESIGNED NESTED-FACTORIAL EXPERIMENT	14
VI.	SUMMARY	18
VII.	REFERENCES	20
VIII.	DISTRIBUTION	21

ABSTRACT

The three basic principles of experimental design: replication, randomization, and local control, are briefly discussed. Advantages of the proper application of the principles are given, and the importance of the role of experimental design is illustrated by a numerical example. Three analyses are performed on the same set of test data. First, an analysis is performed on a set of empirical data to salvage as much information as possible from the data of the uncontrolled experiment. Secondly, some (but not all) information is obtained about the conduct of the experiment, necessary assumptions are then made, and an analysis of variance is performed. Finally, the data are properly analyzed by performing a nested-factorial analysis of variance as dictated by the design and the actual conduct of the experiment. The validity and the amount of information resulting from the three analyses are then compared.

I. INTRODUCTION

The three general phases of experimentation are - the design phase, the execution phase, and the analysis phase. The design phase involves the complete set of actions taken prior to the conduct of the experiment; the execution phase refers to the actual conduct of the experiment; and the analysis phase includes data reduction, numerical computations, and interpretation of results. The importance of the design phase of experimentation cannot be overemphasized because of the dependence of the analysis phase upon the design phase. That is, the basis of the interpretation of experimental data is the analysis, but the analysis is dictated by the experimental design. Addelman (1969) recently expressed concern over the fact that far more emphasis has been placed on analysis than on design in the literature on the design and analysis of experiments.

"Designing" an experiment simply means "planning" an experiment so that information will be collected which is relevant to the problem under investigation. Naturally, this planning phase is the time to ensure that the appropriate quantity and quality of data will be obtained in a manner which permits the proper application of inductive statistical methodology and, consequently, an objective analysis leading to valid inferences with respect to the stated problem.

The objective of any experimental design is to provide the maximum amount of information at a minimum cost. Consequently, experimental design is concerned with both statistical efficiency and resource economy. Both features should be present in any scientific investigation.

II. BASIC PRINCIPLES OF EXPERIMENTAL DESIGN

The three basic principles of experimental design, replication, randomization, and local control, are well summarized by Chew (1958). Replication serves a dual purpose. It makes a statistical test of significance possible (by providing a valid estimate of experimental error), and it improves the precision of the estimated effects of the factors under investigation. While replication makes a test of significance possible, randomization makes the test valid by eliminating bias and by making it appropriate to analyze the data as though the errors were independent. Errors from experimental units adjacent in time or space tend to be correlated, but the randomization gives any two "treatments" being compared an equal chance of being adjacent.

For example, suppose two tank types (A and B) are to be compared with four test runs each. The order of the eight runs should be completely randomized. From an operational viewpoint, a design like AAAABBBB or BBBBAAAA might be more convenient. However, both of these designs are poor because tank effect and time effect are "confounded." The weather, visibility, ground condition, crew fatigue, etc., may be quite different during the first four runs than they are during the last four runs. Consequently, that difference attributed to tank effect may be grossly inflated due to the presence of other indistinguishable effects because of improper randomization. Proper randomization would guard against continually favoring or handicapping either tank type. Cochran and Cox

(1957) describe randomization as "insurance against disturbances that may or may not occur and that may or may not be serious if they do occur."

Replication and randomization make a valid test of significance possible. Local control then makes the test more sensitive (or powerful) by reducing experimental error. That is, local control makes the experimental design more efficient through the use of such features as balancing, blocking, and grouping of the experimental units.

The following partial list contains areas of concern during the design phase:

1. Choice of response or dependent variable.
2. Identification of existing independent variables (factors) involved.
3. Identification of controllable and uncontrollable factors.
4. Selection of controllable factors to be varied.
5. Identification of levels of these controllable factors.
6. Identification of qualitative and quantitative factors.
7. Identify factors having fixed levels and those having random levels.
8. Relationship of factors (crossed or nested).
9. Restriction upon randomization.
10. Method of randomization.
11. Order of experimentation.
12. Formulation of hypotheses.

A "check list" can prove most helpful in assuring that nothing has been overlooked during the design phase. Such a check list is provided in Ostle (1963 and 1967).

Ultimately in the design phase, a mathematical model is hypothesized for the relationship of the dependent variable to the independent variables. That is, the response variable is expressed as a function of the independent variables. This hypothesized model, along with all the necessary assumptions concerning the model, provides the basis for a statistical analysis which is performed on the experimental data. An outline of a proposed statistical analysis at this point provides an excellent opportunity for ensuring that the analysis will, in fact, accomplish the objectives of the experiment.

Many advantages, both direct and indirect, can result if full use is made of the principles of experimental design. A partial list of the advantages of statistically designed experiments is as follows:

1. The statement of experimental objectives is usually developed more completely.

2. The required coordination between the analyst(s) and the experimenter(s) facilitates the analysis, the interpretation of results, and the drawing of conclusions.

3. Attention is focused on interrelationships among the variables under investigation.

4. Sources of variability are identified and measured with increased accuracy and precision.

5. The number of experimental units required to achieve a stated objective can generally be accurately estimated and often reduced.

6. An estimate of experimental error is usually obtained.

7. A greater quantity of usable data is obtained for each dollar expended.

8. Analysis can be improved by eliminating incorrect analysis resulting from a lack of understanding how the experiment was conducted.

9. Cooperation can be improved between groups not in complete contact with one another during the execution and the analysis phases.

10. The invalid extrapolation of data beyond the range of experimental conditions can be avoided.

In the sections which follow, the role of experimental design is illustrated. In the first example, a "salvage operation" is illustrated for empirical data from an uncontrolled experiment. Then the data are analyzed as if it were obtained from a designed experiment. But, a wrong model is employed, and the underlying assumptions are ignored. Finally, the data are properly analyzed as dictated by the design and conduct of the experiment. The quality and quantity of the information obtained from the three analyses are discussed.

III. "SALVAGE" OPERATIONS ON AN UNCONTROLLED EXPERIMENT

Consider an investigation concerning evaluation of concepts, doctrine, and organization of field artillery. Suppose a specific facet of the test is an evaluation of a newly proposed loading method. Cost of the proposed loading method is the same as the currently employed method; therefore, the proposed method will be recommended if the loading rate of the new method is significantly faster than the old method.

A field test was conducted and 18 trials were performed for each loading method. The number of rounds loaded were recorded for each trial. The average number of rounds loaded per minute for each trial is shown below.

TABLE I

FIELD ARTILLERY DATA*

Trial Number	New Method (Rounds per Min.)	Old Method (Rounds per Min.)
1	20.2	14.2
2	26.2	18.0
3	23.8	12.5
4	22.0	14.1
5	22.6	14.0
6	22.9	13.7
7	23.1	14.1
8	22.9	12.2
9	21.8	12.7
10	24.1	16.2
11	26.9	19.1
12	24.9	15.4
13	23.5	16.1
14	24.6	18.1
15	25.0	16.0
16	22.9	16.1
17	23.7	13.8
18	23.5	15.1

* From Hicks (1964), Fundamental Concepts in the Design of Experiments, p. 172.

With the above information, the two loading methods are to be evaluated. Such an evaluation might take the form of a comparative analysis of the number of rounds loaded per minute by the two methods. The analysis should begin with a study of the two sample distributions. Suppose no information exists suggesting that the parent populations of the two samples are non-normal. An appropriate test such as the Kolmogorov Test

could be performed to test the null hypothesis that the parent population of each of the two samples is a normal distribution with mean and variance equal to those of the corresponding sample distribution. Rather than illustrate the Kolmogorov Test, which is not the purpose of this paper, normality of the parent populations will be assumed (or assumed to have been tested and not rejected).

Having established normality of the two populations, equality of the two population variances (σ_1^2 and σ_2^2) must be investigated. That is, the null hypothesis H_o: $\sigma_1^2 = \sigma_2^2$ is tested against the alternative hypothesis H_a: $\sigma_1^2 \neq \sigma_2^2$. The appropriate test for testing equality of the population variances is the variance ratio s_1^2/s_2^2 which is F distributed with degrees of freedom equal to the degrees of freedom of s_1^2 and s_2^2, where s_1^2 and s_2^2 are sample estimates of the population parameters, σ_1^2 and σ_2^2. Notationally,

$$\frac{s_1^2}{s_2^2} \sim F(n_1-1, n_2-1),$$

where n_1 is the sample size of the new method and n_2 is the sample size of the old method ($n_1 = n_2 = 18$). The familiar statistics ($s_j^2; j = 1,2$) estimating the population variances are:

$$s_j^2 = \sum_{i=1}^{n_j} (x_{ji}-\overline{x}_j)^2/(n_j-1); \; j = 1,2$$

where $$\overline{x}_j = \sum_{i=1}^{n_j} x_{ji}/n_j; \; j = 1,2$$

and x_{1i} and x_{2i} refer to sample values from the new and the old method, respectively. The above four computed statistics are:

$$\bar{x}_1 = 23.59 \qquad \bar{x}_2 = 15.08$$

$$s_1^2 = 2.512 \qquad s_2^2 = 3.889$$

Because the alternative hypothesis ($\sigma_1^2 \neq \sigma_2^2$) contains both inequalities, $\sigma_1^2 < \sigma_2^2$ and $\sigma_1^2 > \sigma_2^2$, the variance ratio test is a two-sided test. That is, the rejection region for the test statistic is:

$$s_1^2/s_2^2 < F(\alpha/2, n_1-1, n_2-1)$$

or

$$s_1^2/s_2^2 > F(1-\alpha/2, n_1-1, n_2-1)$$

Arbitrarily selecting $\alpha = 0.05$ as the level of significance for illustrative purposes, $s_1^2/s_2^2 = 0.646$ is compared with the critical F-values, $F(0.025,17,17) = 0.37$ and $F(0.975,17,17) = 2.67$, and is seen to be well within the acceptance region. The best estimate, then, of the common but unknown population variance (σ^2) is the pooled variance,

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

$$= 3.022$$

Having established that the two parent normal populations have

a common variance, the two-sample t-test is applicable for testing the inequality of the two population means (μ_1 and μ_2). Unlike the above two-sided F-test, the appropriate t-test is a one-sided test because we are interested in replacing the old method with the new method only if it is significantly faster (rounds loaded per minute) than the old method. That is, we consider the serious error to be wrongly concluding that the new method is faster when, in fact, it is not (type II error is ignored). Consequently, the null hypothesis H_0: $\mu_1 \leq \mu_2$ is tested against the alternative hypothesis H_a: $\mu_1 > \mu_2$. The computed test statistic (t_c) is:

$$t_c = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

with n_1+n_2-2 degrees of freedom. If $t_c \geq t(1-\alpha, n_1+n_2-2)$, reject the null hypothesis at the α-level of significance; otherwise do not reject the null hypothesis. Again, $\alpha = 0.05$ is arbitrarily selected for illustration. The tabulated critical value is $t(0.95,34) \approx 1.69$, and the computed t-statistic is $t_c = 14.67$. Therefore, the null hypothesis is rejected at the 0.05-level of significance, i.e., the number of rounds loaded per minute by the new method is not equal to or less than the number of rounds loaded per minute by the old method.

A cursory examination of the above analysis might suggest that the analysis accomplished the objective of the investigation. However, further examination quickly reveals that the analysis leaves many questions unanswered. These questions concern, but are not limited to, the order of

testing the loading methods, the number of loading teams used in the test, and the physique and level of training of the men participating in the test. Unfortunately, the available information about the test is inadequate to answer questions in the above cited areas. Because of incomplete information concerning the conduct of the test, the above analysis is, essentially, the extent of the statistical analysis which can be performed and have validity.

IV. "DESIGNED" AFTER EXECUTION

Suppose a more thorough statistical analysis is desired. A query into the conduct of the test revealed that in addition to the two loading methods, two other factors were controlled. The loading teams consisted of three teams for each of three physique classification groups, and each method, group, team combination was tested twice. The same test data given in TABLE I above is retabulated in the following three-way table illustrating the three controlled factors.

TABLE II

THREE-WAY DATA ARRAY

	Group								
	1			2			3		
Teams	1	2	3	1	2	3	1	2	3
Method 1	20.2	26.2	23.8	22.0	22.6	22.9	2[illegible].1	22.9	21.8
	24.1	26.9	24.9	23.5	24.6	25.0	22.9	23.7	23.5
Method 2	14.2	18.0	12.5	14.1	14.0	13.7	14.1	12.2	12.7
	16.2	19.1	15.4	16.1	18.1	16.0	16.1	13.8	15.1

The test data is now analyzed by employing the analysis of variance (ANOVA) procedure. ANOVA is a method of partitioning the total variability of a response variable into component parts associated with the controlled factors under investigation and the uncontrolled random error. Methods, Groups, and Teams are termed "factors," and the classifications of the factors are termed "levels." That is, the number of levels for Methods, Groups, and Teams are two, three, and three, respectively. The ANOVA model used is:

$$y_{\alpha\beta\gamma\rho} = \mu + A_{\alpha} + B_{\beta} + C_{\gamma} + AB_{\alpha\beta} + AC_{\alpha\gamma} + BC_{\beta\gamma} + ABC_{\alpha\beta\gamma} + R_{\alpha\beta\gamma\rho}; \quad \begin{aligned} \alpha &= 1,2 \\ \beta &= 1,2,3 \\ \gamma &= 1,2,3 \\ \rho &= 1,2. \end{aligned}$$

where y is the response variable; μ is the true mean effect; A, B, and C are the Method, Group, and Team effects, respectively; AB, AC, and BC are the two-factor interaction effects between the factors; ABC is the three-factor interaction effect; and R is the random error effect.

Neither the assumptions of the model nor the computational procedures of the ANOVA are discussed. This is not to be construed that assumptions of the model are not important; the assumptions are very important as will be illustrated later. The computational results are illustrated and the inferences from the analysis are discussed. The usual ANOVA table of the computational results is below:

TABLE III

ANOVA FOR FACTORIAL EXPERIMENT

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio
A	1	651.95	651.95	282.23***
B	2	16.05	8.02	3.47
C	2	12.77	6.38	2.76
AB	2	1.19	0.60	0.26
AC	2	5.56	2.78	1.20
BC	4	26.49	6.62	2.87
ABC	4	5.16	1.29	0.56
Error	18	41.59	2.31	
Total	35	760.76		

*** - Significant at the 0.001-level of significance

NOTE - A fixed effects model is assumed.

The previously mentioned partitioning of the total variability is illustrated in the Sum of Squares Column in the table. Note that the Total Sum of Squares (760.76) is partitioned into eight component parts. For example, of the total variability (760.76) within the 36 data values, 651.95 of it is attributed to Methods (represented by factor A). A visual inspection of the Sum of Squares Column suggests that Methods is the predominant source of variation of the average number of rounds loaded per minute during the test.

The actual testing of the various hypotheses concerning the factorial effects of the model is accomplished by a comparison of the values in the

F-Ratio Column with appropriate critical F-values previously discussed. For example, testing at the 0.05-level of significance for illustration, 282.23 is compared with F(0.95,1,18) = 4.41 and is found to be highly significant. Continuing the testing as illustrated, none of the remaining F-ratios are found to be significant.

A first impression of the above analysis of variance might be that the analysis is thorough and complete, and the objective of the investigation has been satisfactorily accomplished. An examination of the analysis, however, reveals that the analysis is not valid. Recall that the design of an experiment dictates the analysis of the experimental data collected.

In the above analysis of variance (TABLE III), all factorial effects were tested against the within error variance. That is, the F-ratios were determined from MS(Factorial Effect)/MS(Error). However, the within error variance is the denominator of all F-ratios only if the ANOVA model is a fixed effects model (or Model I). An examination of the methods of selecting the levels of each of the three factors indicated that all three factors were not fixed; the teams were randomly chosen. Therefore, the ANOVA model is a mixed model (or Model III), i.e., the model contains both fixed and random factors; factors A and B are fixed, and factor C is random. To determine the proper F-ratios for testing the factorial effects, the expected mean squares (EMS's) must be determined because the F-ratios are derived from the EMS's.

The usual rules available in the literature may be employed to determine the EMS's (See, for example, Bennett and Franklin (1954), Hicks (1964), or Johnson and Leone, Vol. II (1964)). The EMS's for the above particular model, however, have already been derived and are available. The correct F-ratios for the particular mixed model are illustrated, for example, in Ostle (1963), Wine (1964), and Beyer (1966) - MS(A)/MS(AC), MS(B)/MS(BC), MS(C)/MS(Error), MS(AB)/MS(ABC), MS(AC)/MS(Error), MS(BC)/MS(Error), MS(ABC)/MS(Error).

V. PROPERLY DESIGNED NESTED-FACTORIAL EXPERIMENT

In addition to the incorrect F-ratios in the above analysis, further investigation into the conduct of the experiment revealed that the ANOVA model in Section IV was incorrect. The three randomly selected teams (Factor C) were not the same teams for all three groups (Factor B). Three different teams were randomly selected for each of the three groups. Therefore, nine teams were actually used in the experiment. The experiment, therefore, was not a factorial experiment; it was a nested-factorial experiment.

To emphasize the nesting feature of the experiment, the data layout of TABLE II is revised. TABLE IV below reflects the nesting of the teams within groups, and TABLE V contains the cell and marginal means of TABLE IV.

TABLE IV

DATA LAYOUT FOR NESTED-FACTORIAL EXPERIMENT

	Group								
	1			2			3		
Teams	1	2	3	4	5	6	7	8	9
Method 1	20.2	26.2	23.8	22.0	22.6	22.9	23.1	22.9	21.8
	24.1	26.9	24.9	23.5	24.6	25.0	22.9	23.7	23.5
Method 2	14.2	18.0	12.5	14.1	14.0	13.7	14.1	12.2	12.7
	16.2	19.1	15.4	16.1	18.1	16.0	16.1	13.8	15.1

TABLE V

CELL AND MARGINAL MEANS

	Group									
	1			2			3			
Teams	1	2	3	4	5	6	7	8	9	
Method 1	22.15	26.55	24.35	22.75	23.60	23.95	23.00	23.30	22.65	23.59
		24.35*			23.43*			22.98*		
Method 2	15.20	18.55	13.95	15.10	16.05	14.85	15.10	13.00	13.90	15.08
		15.90*			15.33*			14.00*		
	18.68	22.55	19.15	18.92	19.82	19.40	19.05	18.15	18.28	
		20.12**			19.38**			18.49**		19.33

* - Method-Group Means

** - Group Means

Knowledge of the relationships (crossed or nested) of the factors under investigation is necessary before the correct analysis of variance model can be specified. However, this alone is not sufficient; knowledge of the

order of experimentation is also necessary. In the field artillery experiment, no blocking was performed; the order of experimentation was completely randomized. Therefore, the correct model for the nested-factorial experiment in a completely randomized design is:

$$y_{\alpha\beta\gamma\rho} = \mu + A_\alpha + B_\beta + AB_{\alpha\beta} + C_{\gamma(\beta)} + AC_{\alpha\gamma(\beta)} + R_{\rho(\alpha\beta\gamma)} \quad ; \quad \begin{array}{l} \alpha = 1,2 \\ \beta = 1,2,3 \\ \gamma = 1,2,3 \\ \rho = 1,2. \end{array}$$

The parentheses in the model denote the nesting of the factors. The factor(s) represented by the subscript(s) not in parentheses is (are) nested within the factor(s) represented by the subscript(s) within parentheses. For example, $C_{\gamma(\beta)}$ denotes that factor C is nested within factor B. Note additionally, from the ANOVA model and from TABLE IV that Method (factor A) and Group (factor B) are crossed, while Method is crossed with Teams (factor C) within Groups.

The sums of squares corresponding to like terms of the correct model and the incorrect model are identical. The sums of squares corresponding to the two unlike terms, $C_{\gamma(\beta)}$ and $AC_{\alpha\gamma(\beta)}$, are

$$SS\left[C_{\gamma(\beta)}\right] = 39.26$$

$$SS\left[AC_{\alpha\gamma(\beta)}\right] = 10.72$$

In order to determine the F-ratios in the analysis of variance, the rules cited in Section III are applied. The resulting expected mean squares and the F-ratios to be performed are illustrated in the following table.

TABLE VI

EXPECTED MEAN SQUARES

Source	Expected Mean Square
A_α	$\sigma^2 + 2\sigma^2_{AC} + 18 \sum_{\alpha=1}^{2} A^2_\alpha$
B_β	$\sigma^2 + 4\sigma^2_C + 12 \sum_{\beta=1}^{3} B^2_\beta/2$
$AB_{\alpha\beta}$	$\sigma^2 + 2\sigma^2_{AC} + 6 \sum_{\alpha=1}^{2} \sum_{\beta=1}^{3} AB^2_{\alpha\beta}/2$
$C_{\gamma(\beta)}$	$\sigma^2 + 4\sigma^2_C$
$AC_{\alpha\gamma(\beta)}$	$\sigma^2 + 2\sigma^2_{AC}$
$R_{\rho(\alpha\beta\gamma)}$	σ^2

The proper analysis of variance for the experiment is given in the following table.

TABLE VII

ANOVA FOR NESTED-FACTORIAL EXPERIMENT

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio
A_α	1	651.95	651.95	364.22***
B_β	2	16.05	8.02	1.23
$AB_{\alpha\beta}$	2	1.19	0.60	0.34
$C_{\gamma(\beta)}$	6	39.26	6.54	2.83*
$AC_{\alpha\gamma(\beta)}$	6	10.72	1.79	0.77
Error	18	4[illegible].59	2.31	
Total	35	760.76		

* - Significant at the 0.05-level of significance

*** - Significant at the 0.001-level of significance

From a comparison of the sum of squares in TABLES III and VI, the sum of the two sum of squares due to C and BC of the incorrect model is found to be the sum of squares due to $C_{\gamma(\beta)}$. Similarly, the two sum of squares due to AC and ABC of the incorrect model are found to be the sum of squares due to $AC_{\alpha\gamma(\beta)}$. That is, what was formerly thought to be Team effect and Group-Team interaction effect is, in fact, Team-within-Group effect. And, what was wrongly thought to be Method-Group interaction effect and Method-Group-Team interaction effect is, in fact, Method-Team-within-Group effect.

Although the F-ratios in TABLE III are incorrect, conclusions concerning the significance of Methods, Groups, and Method-Group interaction are not changed drastically. However, in the correct analysis a difference between Teams-with-Groups is significant at the 0.05-level of significance which was not detected in the incorrect analysis. This team difference suggests further team investigation is required. That is, knowledge of the reason for the team difference would be desirable. Further investigation of teams within the three physique groups might attribute the difference to such things as extent of individual experience, time since training, duration of training, location of training, method of training, etc.

VI. SUMMARY

The above three analyses (t-test, factorial ANOVA, and nested-factorial ANOVA) performed on the same set of data illustrate the important role of design in experimental investigations. The t-test was a valid analysis, but the analysis did not extract all the available information from the

experimental data. More information was available from the data than the analysis revealed. The factorial ANOVA provided more information than the t-test. However, because the ANOVA model and some of the underlying assumptions were incorrect, the analysis was invalid. Consequently, the additionally obtained information was incorrect. The nested-factorial ANOVA, on the other hand, did yield valid information as well as more information. Because the ANOVA model and the assumptions concerning the model were determined from a complete description (to the extent possible) of the actual conduct of the experiment, the resulting analysis was valid. Further, the additionally obtained information concerning team differences was bonus information that may be utilized when planning future experiments.

The above analyses also illustrate the importance of planning before execution of the experiment. Deciding after the conduct of an experiment is not the time to design an experiment. An experiment must be designed before its execution. Only then can assurance be made that all factors have been properly considered, levels of the factors have been properly chosen, and the order of experimentation has been properly determined.

The value of statistically designed experiments is evident and should always be sought. Experience has shown that the return for the effort spent in designing an experiment far outweighs the expense. In short, the importance of the design phase of an experimental investigation cannot be overemphasized.

VII. REFERENCES

1. Addelman, S. (1969), The Generalized Randomized Block Design, The American Statistician, Vol. 23, No. 4, pp. 35-36.

2. Bennett, C. A. and Franklin, N. L. (1954), Statistical Analysis in Chemistry and Chemical Industry, John Wiley and Sons, Inc., New York.

3. Beyer, W. H. (Ed.), (1966), Handbook of Tables for Probability and Statistics, The Chemical Rubber Co., Ohio.

4. Chew, V. (Ed.), (1958), Experimental Designs in Industry, John Wiley and Sons, Inc., New York.

5. Cochran, W. G. and Cox, G. M. (1957), Experimental Design, 2nd Ed., John Wiley and Sons, Inc., New York.

6. Hicks, C. R. (1964), Fundamental Concepts in the Design of Experiments, Holt, Rinehart & Winston, New York.

7. Johnson, N. L. and Leone, F. C. (1964), Statistics and Experimental Design in Engineering and the Physical Sciences, Volume II, John Wiley and Sons, Inc., New York.

8. Ostle, B. (1963), Statistics in Research, 2nd Ed., Iowa State University Press, Ames, Iowa.

9. Ostle, B. (1967), Industry Use of Statistical Test Design, Industrial Quality Control, Vol. 24, No. 1, pp. 24-34.

10. Wine, R. L. (1964), Statistics for Scientists and Engineers, Prentice-Hall, Inc., New Jersey.